

25 E 000 4603030

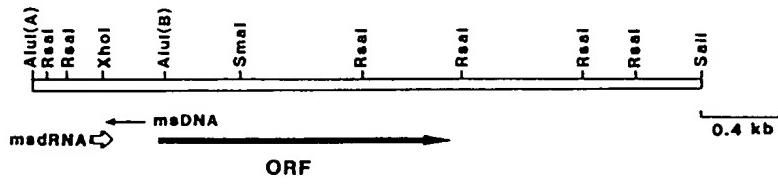


FIGURE 1

**FIGURE 2**

D  
E  
E  
B  
O  
B  
B  
B  
B  
B  
B  
B

HIV	RT	VKLXPGMDCPKVQ WPLTEEKIALVEICTEMEKECKISKICPENFYNTPVFAIKKDSTKWR	
HTLV1	RT	RPWARTPPKAPRNQ PVPFKPERLQALQHILVRKALEACHIPEPYTC PGNNPVFPVKKA NGTUR	
Ec-67	RT	NVLYRICSDNQYTQFTIPKKGKGVRTISAPTDRL KDIQRRICDLLSDCRDEIFAIRKI SNMYS	
Mx-162	RT	AFHREVDATATHYVSWTIPKRDGSKRTITSPKEL KAAQR WVLS NVV ERLP VHCAA	
		o oooo ooo o x xx oo oo x	
HIV	RT	KLVDFRELNRKTQDFWEVQLCIPHAGLKKK KSUTVLDVGDAYFSVPLDEDFRKYT	A
HTLV1	RT	FIHDLRATNSLTIDLSSSPCPDLSLPTTLAHLQTIDLRAFFQIPLPKQFQPYF	A
Ec-67	RT	FCFE RGKSIILNAYKHRCKQIILNIDLKDFESFNFCRVRG YFLS NQDF	L
Mx-162	RT	HGFV AGRSILTNALAHQCADVVVKVDLKDFPSVTWRRVKGLRKGLREGTSTLLSLLSTEAP	
		oo o oo x o o oooooo xxx oo o x x	
HIV	RT	FTIP SINNETPGIRYQYNVLPOGWKGSPAIFQS SMTKILEPFKKQNPDIVIYQYHDDLYVC	
HTLV1	RT	FTVP QQCNYCPCTRYAWKVLPOQFKNSPTLFEM QLAHILQPIRQAFPQCTILQYHDDLLLA	
Ec-67	RT	LN PVVATTLAKAACYN CTPCGSPCSPIISNLICNIHMDRLAKLAKKY GCTYSRYADDITI	
Mx-162	RT	REAVQFRCKLLHVAKGP RALPOGAPTSPOITNALCLKLDRKRLSALAKRL GFTYTRYADDLTF	
		o o ooo o o x x oo oo x o oooooo x	
HIV	RT	S DLEIGQHRTKIEELRQHLLRWGLLTP DKKHQKEP PFLWMCYELHPDKWTQPIVLPE KD	
HTLV1	RT	S PSHEDELLSEATHMASLISHGLPVS ENKTQQTPGTIKFLCQIISPNNHLTYDAVPTVPI RS	
Ec-67	RT	STNKNTFPLEATVQPEGVVLCKVLVKEIENSFEINDSKTRLYTKTSRQEVT CLTVNRIVNID	
Mx-162	RT	SWTAKAKQPKPRRTQRPVAVLLSRVQEVEAEGRVHPDKTRVARKCTRQRT GLVVNAACKDA	
		* o o o o o x o oo ooo o oo oo oo xx	
HIV	RT	SWTVNDIQKLVGKLNWASQIYP	
HTLV1	RT	RWALPELQALLGEIQWVSKGTP	
Ec-67	RT	RCYYKKTRALAHALYRTCE YK	
Mx-162	RT	PAARVPRDVVRQLRAAIHN RK	
		x o	

FIGURE 3

**A**

Mx-162	18	PTPELTAPSSDAAAKREARRLAHEALLVRAKAIDEAGGADDNVQAQLVSKGLAVEDLD-FSSASEKDCKKA-WKEKK	91
Mo-MLV	1070	PDPDMTRVTNSPSIQAHLQALYLVQHEVW-RPL-AAAYQEQQ-LDRPVVPHPYRVDTVWVRRHQTKNLEPRWKGPY	1142

<sup>o</sup><sup>o</sup><sup>o</sup><sup>o</sup><sup>o</sup><sup>o</sup><sup>o</sup><sup>o</sup><sup>o</sup>**B**

Mx-162	92	KAEATERRALKRQAHEAW-KATHVGHLGAGVHWAEDRL	128
Mo-MLV	1143	TVLLTTPTALKVDGIAAWIHAAHVKAADPGGG-PSSRL	1179

<sup>o</sup><sup>o</sup><sup>o</sup><sup>o</sup><sup>o</sup><sup>o</sup><sup>o</sup><sup>o</sup><sup>o</sup>

Mx-162 411 GKDAPAAPVPRDVVRQLRAAIHNRKKGKGREGESLEQLKGMAAFIHMTD-PAKGRAF-LAQLTELESTASAAPQAE 485

HIV 396 GKEGHSAARQCR-APR--RQGC--WKCGKPGHIMTNCPD-R-QAGFLGLGPWGKKPRNFPVAQVPQ-GLTPTAPP 461

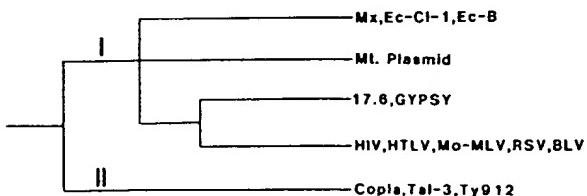
<sup>o</sup><sup>o</sup><sup>o</sup><sup>o</sup><sup>o</sup><sup>o</sup><sup>o</sup><sup>o</sup><sup>o</sup>

Figure 4. Sequence Similarity of the msDNA-Mx162 Reverse Transcriptase with Other Retroelements  
(A) Sequence similarity of the region from residues 18 to 128 of the msDNA Mx162 RT (see Figure 2) with a carboxy-terminal region of integration protein of Moloney murine leukemia virus (M-MLV) (residues 1070 to 1179; Shinnick et al., 1981)  
(B) Comparison of the sequence from residues 411 to 485 of the msDNA-Mx162 RT (see Figure 2) with the sequence from residues 396 to 461 of the gag protein of human immunodeficiency virus (HIV; Ralner et al., 1985).

FIGURE 4

**A**

Mx-162	304 GP-RALPQGAPTPSGITNALCLKLDKRISALAKRL-GFTYTRYADELTF-SWTKAQPKPRRTQRPPVAVL	371
Ec-67	159 YN-GTLPQGSPCSPPIISNLICNIMDRILAKLAKKY-GCTYSRYADDITI-STNONTFPLEMATVQPEGVVL	226
Ec-86	130 YK-NLLPQGAPSSPKLANLICSKLDYRIQQYAGSR-GLIYTRYADELTL-SAQSMKKVVVKARDFLFSIIPS	197
HIV	311 YQYNVLPQGWKGSPAIFQS---SMTKILEPFKKQNPDIVIYQYMDILYVGGS-DLEIGQHRTKIEELRQHLL	377
HTLV1	150 YAWKVLPQGFKNSPTLFEM---QLAHILQPIRQAFPQCTILQYMDILLLAS--PSHEDLLLSEATHMASLI	215
Mo-MLV	303 LTWTRLPLQGFKNSPTLFDE---ALHRDLADFRHQPDLLLQYVDDILLAA-TSELDCCQG-TRALL-QTL	367
RSV	141 FQWKVLPQGMTCSPTICQL---VVGQVLEPLRLXHPSLCMLHYMDILLLAA--SSHDGLEAACEEVI-STL	205
BLV	122 FAWRVLPQGFINSPALFER---ALQEPLRQVSAAFSQSLSLVYMDILLYAS--PTEEQRSQCYQALA-ARL	186
Mt. plasmid	288 IATNGVPQGASTSGLATYNVL-----ELFLRY--DELIMYADGGIL-CRQDPSTPDFSVEEAGVVQEP	348
17.6	339 YEYLRLMPFGLRNAP-ATFQRCMN-DI---LRPLLNNKHC-LVYLDLIIIVFS-TSLDEHLQSLGLVFE--KL	399
GYPSY	284 YEFCRLPFGLRNASSIFOR---ALDDV---LREQI-GKICVYVYDDVIIFS--ENESDHVRHIDTVLK-CL	344
Copia	1032 CKLNKAIYGLKQAARCWF'R-CIYI---LDKGHNINENIYV-LVYDDEVIAT--GDMTRMNNFKRYLME-KF	1112
Tal-3	990 CLLKKSLYGLKQSPRQWNA-CVYV-KQVSE-QEHLYL---LLVVDDEMLIAG--KSKSEINKVKEQLSM-EF	1069
Ty912	948 IRLKKSLYELKQS-GANWYE--EVRC-WSCVFKHSQV-T1CLFVDDAVLFS--KNLNSNKRIIEKLKM-QY	1023

**B****FIGURE 5**

S 1 2 3 4

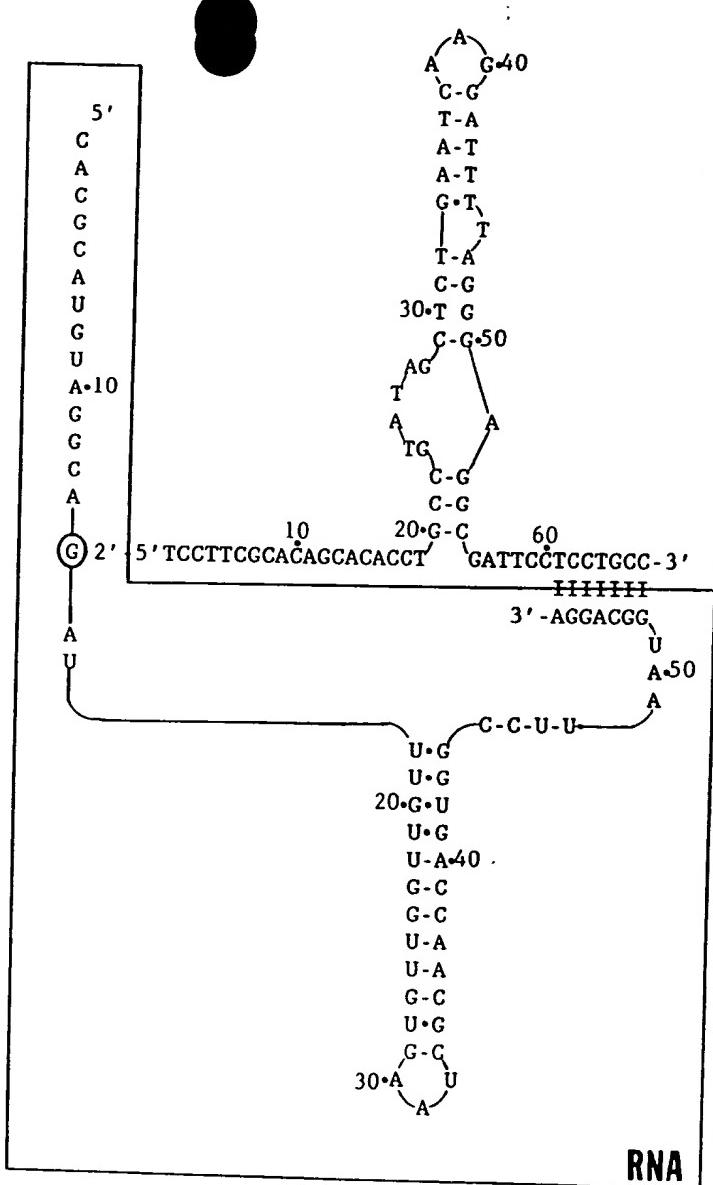


FIGURE 6

FIGURE 7

2000-08-09

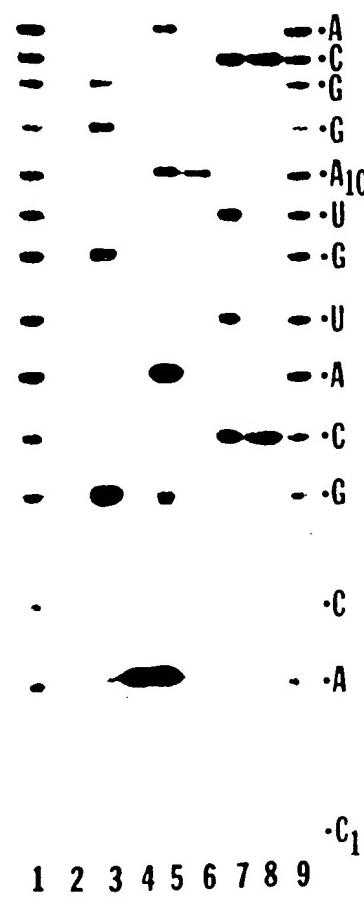
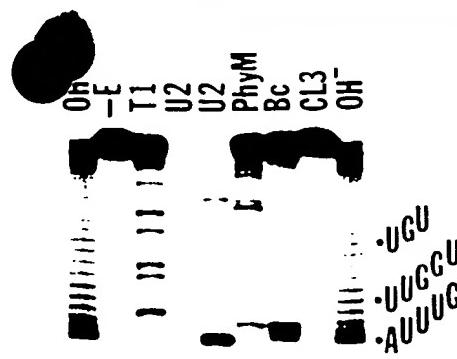
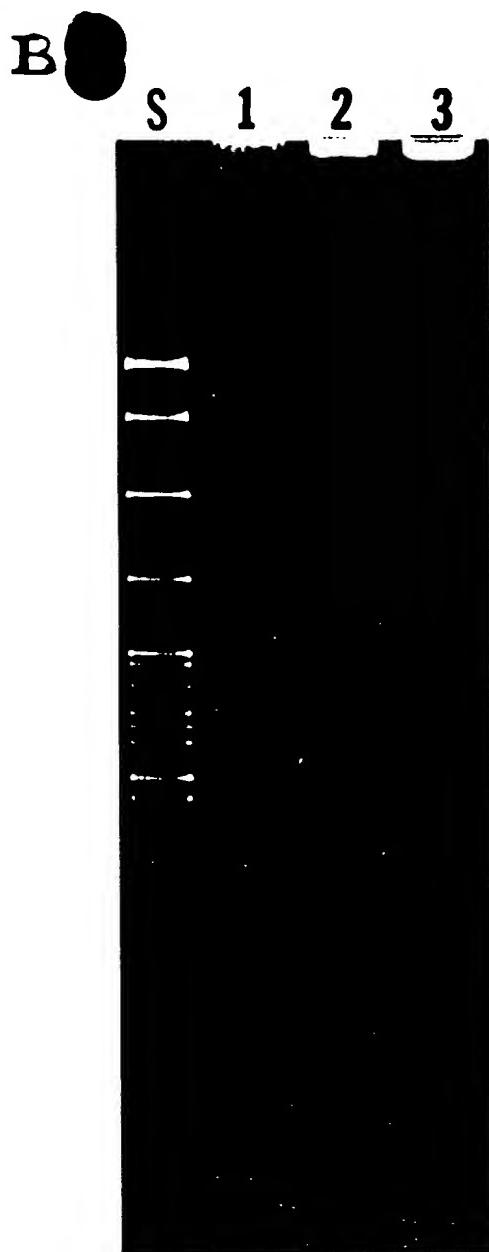
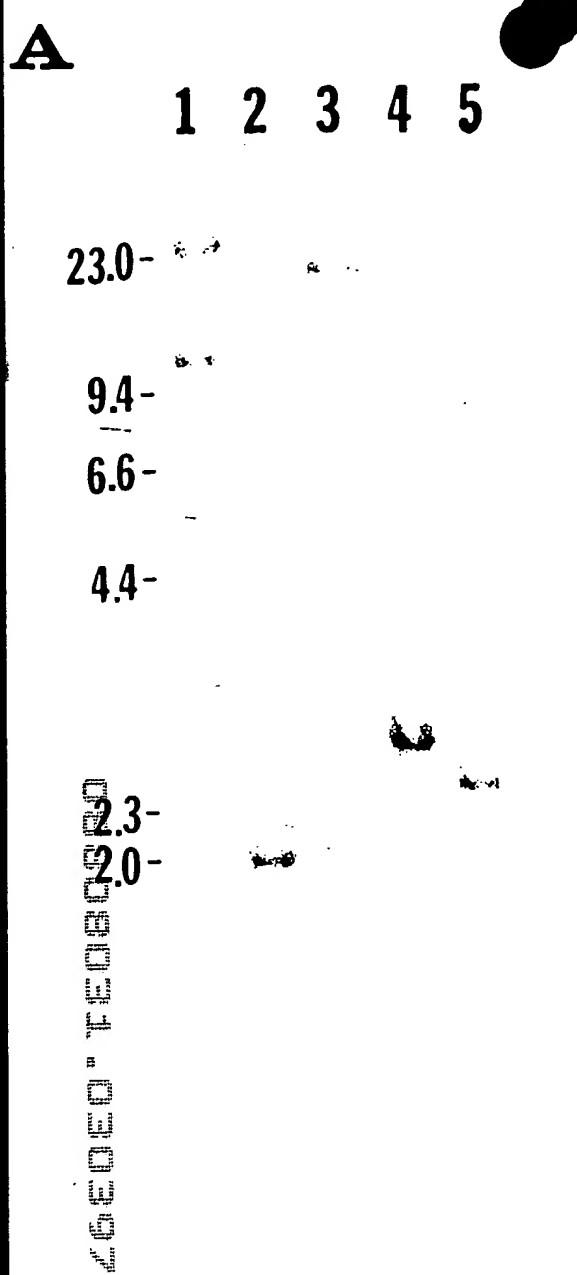


FIGURE 8



FIGU-

**FIGURE 9**

0 8 0 8 0 3 1 - 0 3 0 3 9

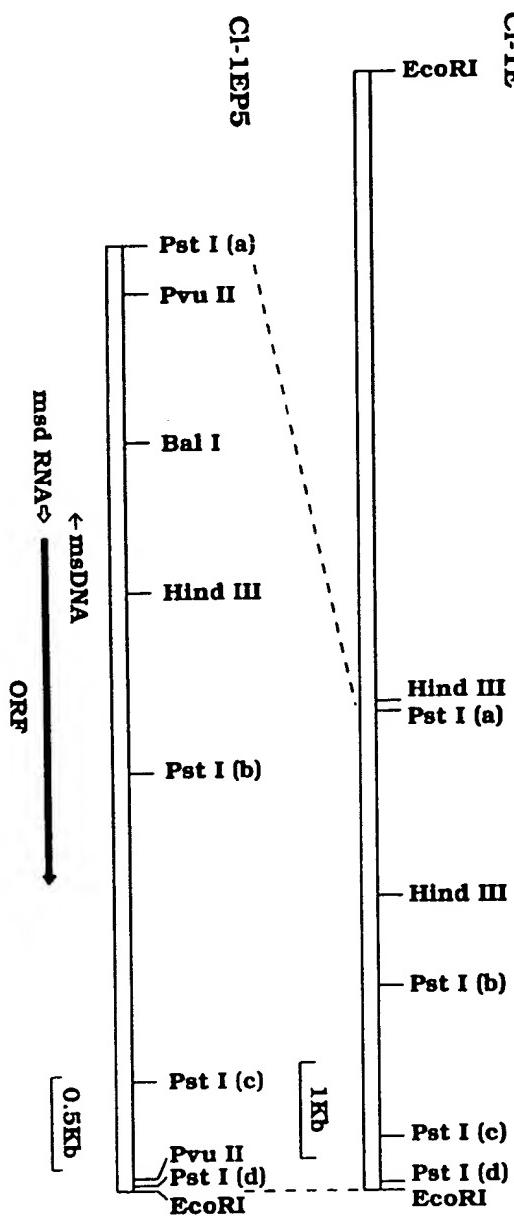


FIGURE 10

### FIGURE 11

V	RT	VKLKPGMDGPKVKQ	WPLTEEKIKALVEICTEMEMEKEGKISKIGPENPYNTPVFAIKKKDSTKWR	239
LV1	RT	RPWARTPPKAPRNQ	PVPFKPERLQALQHLVRKALEAGHIEPYTG	PGNNPVFPVKKA NGTWR 75
DNA	RT	NVLYRIGSDNQYTQFTI	KPKKGKVRTISAPTDRL	KDIQRICDLSDCRDEIFAIRKI SNNYS 94 + o      ●      ● +      +      +      +      +      +
V	RT	KLVDRELNKRTQDFWEVQLGIPHAGLKKK	KSVTVLDVGDAYFSVPLDEDFRKYTAFTIP SI	302
LV1	RT	FIHDLRATNSLTIDLSSSPGPPDSSLPTTLAHLQTIDLRDAFFQIPLPKQFQPYFAFTVP QQ	139	
DNA	RT	FGFE RGKSIILNAYKHRGKQIILNIDLKDFFESFNFGGRVRG YFLS NQDF LLN PVVA 150 o      ●      +      +      + o      +      +      + ●		
V	RT	NNETPGIRYQYNVLPGWKGSPAIFQS	SMTKILEPFKKQNPDIVIYQYMDDLYVGS DLEIG 363	
LV1	RT	CNYGPGTRYAWKVLPGFKNSPNLFEM	QLAHILQPIRQAFPQCTILOQYMDDILLAS PSHE 199	
DNA	RT	TTLAKAACYN GTLPQGSPCSPIISNLICNIIDMRLAKLAKKY GCTYSRVADDITI STNKNTF 212 ●      ●●●      ●● +      +      o +      oo      ●      ●●●      ●		
V	RT	QHRTKIEELRQHLLRWGLTTP	DKKHQKEP PFLWMGYELHPDKWTVQPIVLP KDSWTVNDI 424	
LV1	RT	DLLLSEATMASLISHGLPVS	ENKTQQTPGTIKFLGQIISPNSHLTYDAVPTVPI RSRWALPEL 262	
DNA	RT	PLEMATVQPEGVVLGKVLVKEIENSGFEINDSKTRLTYKTSRQEVT GLTVNRIVNIDRCYYKKT 276 o      +      oo      +      o +      ●      +      + o		
V	RT	QKLVGKLNWASQIYPGIK	VRQLCKLLRGTKALTEVIPLTEEAELAENREILKEPVHGYYD 487	
LV1	RT	QALLGEIQWVSKGPTLRLQPLHSLYCALQRHTDPRDQIYLNPQVQSLVQLRQALSQNCRSRLVQ 327		
DNA	RT	RALAHALYRTGE YKVPDE NGV LVSGGLDKLEGMFIDQVDFKFNNIKKKLNQ PDRYVL 335 oo      +      +      +      + o +      oo      ●      o+o		
V	RT	PSKDLIA EIQQGQGQWTYQIYQE	PFKNLKTGKYARMRGAHTNDVKQLTEAVQKITT 544	
LV1	RT	TLPLLGAIMLTLGTTTVFQSKEQWPLVWLHAPLPHTSQCPWGQLLASAVLLDKYTLQSY GL	391	
DNA	RT	TNATLHGFKLKL NAREKAY SKFIY YKFFHGNTCPTIITEGKTDRIYLKAALHSLET SYPEL 396 o      ●      oo      + oo      + o      o +      + + o +      oo o		
V	RT	ESIVIWGKTPFKLPIQKETWETWWTEYWQATWI	PE WEFV NTPPL VKLWYQ 595	
LV1	RT	LCQTIIHHNISTQTFNQFIQTSDHPSPVILLHHSHRFKNLGAQTGELWNTFLKTAAPLAPVKALMP 456		
DNA	RT	FREKTDSKKKEINLNIFKSNEKTKYFLDLGGTADLKKFVERYKNNYASYYGSV PKQPVIMVLD 460 +      +      oo +      o      o      o      +      o+o      ●      oo		
V	RT	LE KEPIV GAETFYVDGAANRETKLKGAGYVTNKGQK VV PLTNTTNQ KTELQAIYLA 652		
LV1	RT	VFTLSP VIINTAPCLFDGSTSRAAYILWDKQILSQRS FP LPPPQHSA Q RAELLGQHGL 516		
DNA	RT	NDTG PSDLLN FLRNKVKSCPDDVTEMRKMKYIHVFYNYIVLTPLSPSGEQTSMEDLFPKDIL 523 o      ●      o +      + + o+o      +o      o+      o      o      o		
V	RT	LQDS GLE VNIVTDSQYAL QIIQA	QPDKSESELVNQIIIEQLIKKEKVYLAWVPAHKG 708	
LV1	RT	SSAR SWR CLNIFLDSKYLYHQLRTLALGTQGRSSQAPFQA LLPRLLSRKVVYLHHVRSHTN 578		
DNA	RT	DIKIDGKKFNKNNDGDSKTEYGKHI FSMR VV RDKKRKIDFKAFCCIFDA 572 +      ●      ●●○ o +      o o      o      o too +		
V	RT	IGGNEQVDKLVSA	722	
LV1	RT	LPDPISRLNALTDA	592	
DNA	RT	IKDIKEHYKMLNS	586 + o      ++	

FIGURE 12

**B**      **A**

**M-MuLV**

**pGB2**

**pCI-1EP5**

**FIGURE 13**

Sa163	165	RWFSEFHREVD TGTHYQTWEI PKRDGG	K	h R h R	AVQRWVLANV VERLPVH	GAAHG
Mx162	167	RWFAFHREVD TATHYVSWTI PKRDGS	h	KR TITSPKPELK AAQRWVLNSV VERLPVH	-----GAAHG	
Mx65	136	RHYSIHRPRE RVRYHVTFAV PKRSGG	VR	LLHAPKRRLK ALQRRMLALL VSKLPVS	-----PQAHG	
Ec67	29	FLTNVLYRIC SDNQYTQFTI PKKGKG	VR	TISAPTDRLK DIQRRCIDLL SDCRDEI FAI RKISNNYSFG		
Ec86	34	VETLRLLIYT ADFRYRIYTV EKKGPEKRM	MR	TIYQPSRELK ALQGWVLRN1 LDKLSSS	-----PFSIG	
Ec73	14	TKGFASEVMR SPEPPKKWDI AKKKGG	MR	TIYHPSSKV LIQYWLNNN FSKLPMH	-----NAAYA	
Ec107	25	IQLRLHALSNH AGRHYRRIIL SKRHGG	QR	LVLADPYLLK TVQRNLKVN LSQFPLS	-----PFATA	
Consensus		Y-h-h KR-GG	K	• • • • • • • •	-hA-G	
			1	2		
Sa163	225	FVAGRSILTN ALAH--QGAD VVVKVDKDF	A F	FPSVTWPRVK GLLRKGGGLPE	NLATLLALLS TEAPREVVRF	
Mx162	227	FVAGRSILTN ALAH--QGAD VVVKVDLKDF	h Dh G Y h	FPSVTWRRVK GLLRKGGGLPE	GTSTLLSLS TEAPREAVQF	
Mx65	196	FVPGRSIKTG AAPH--VGRR VVLKLDLKDF	h	FPSVTFAVRV GLLKALGYGY	PVAATLAVLM TESERQPVEL	
Ec67	97	FERGKSIILN AYKH--RGKQ IILNIDLKDF	h	FESFNFGRVRV GFYLSNQDFL	LNPVVATTLA -----	
Ec86	96	FEKHOSILNN ATPH--IGAN FVILNIDLEDF	h	FPSLTANKVF GVFSHLGYNR LISSVLT-----		
Ec73	74	FVKNRSIKSN ALLHAESKNK YYVKIDLKDF	h	FPSIKFTDFF YAFTRYRDR EFTTEYDDEL LQLIKT-----		
Ec107	85	YRPGCPIVSN AQPH--COQP QILKLDIENF	h	FDSISWLQVW RVFRQAQLPR NVVTMLT-----		
Consensus		F--GRSIhpN A--H -G-- hhKhDhKDF	• • • • • • • •	•	Shh-----	
		K	K		T	
			3			
Sa163	293	RGETIYVAKG PRALPQGAPT SPALTNALCL	hPOG pp hh h	RLDKRSLALS	n Y DDhh	
Mx162	285	PRELIHVAKG PRALPQGAPT SPGITNALCL	h	KLDKRLSALA	TRYADELTFS WRRAKKSROK	
Mx65	264	EGILEHVPVG PRVCVQGAPT SPAICNAVLL	h	RLDRRLAGLA	--KRLGFTY TRYADELTFS WTKAKOPKPR	
Ec67	155	KAACY NGTLPOQGSPC SPIISNLICN	h	IMDMRLAKLA	--RRYGYTY TRYADELTFS GDDVTA-----	
Ec86	151	KICCY KNLLPQGAPS SPKLANLICS	h	KLDYRIOQGYA	--KKYGCTY SRYADDTIS TNKNTFPLEM	
Ec73	140	ICFIS DSTLPIGFPT SPLIANFVAR	h	ELDEKLTKL NAIDKLNATY	--GSRGFLP TRYADELTLS AQSMKK-----	
Ec107	140	WICCY NDALPQGAPT SPAISNLVMR	h	RFDERIGEWG QCARGITY TRYACDDMTFS TNMKGA-----		
Consensus		h--- --hLPQGAPT SP-h-Nhh--	• • • • • • • •	pp-GhTY TRYADDThS -pp--	• • • • • • • •	
		KLDpRL--h-R				
			4	5		
Sa163	360	ELPLADAPVA LLALARVKGV1 EAEGFTLHPD	Gh h c K h	KTRVORKR--G	hLG h	
Mx162	362	--RTQRPVVA VLLSRVQEVV EAEGFRVHPD	h	KTRVARKR--G	SRQRVTGLVV	
Mx65	327	--LE RVRALAARYV QEEGFEVNRE	h	KTRVORR--G	TRQRVTGLVV	
Ec67	217	--ATVQPEGV VLGVKLVKEI	h	KTRLTYK--T	GAQRVTGTV	
Ec86	209	ENSGFEINDS KTRLTYK--T	h	SRQEVTLGLTV	366	
Ec73	201	SV KARDFLFSII PSEGGLVINSK	h	SQRKVTLGVI	262	
Ec107	198	KTC1ISGP--R	h	GSIVVTGLKV	248	
Consensus		QVKNKVCGLL AELGLSLNKR KGCLIAA--C	•	KRQQVTGIVV	243	
		--h---h-phh p-pGhphppp KT-h--p	• • • •	ppQpVTGL-V	235	
			6	7		

FIGURE 14

卷之三

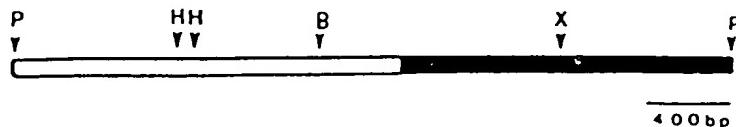
**FIGURE 15**

## FIGURE 16

TTTGAGAAG CGCCATACCA AACAGGGAT ACAGACCAAC CTGACGCTGA AAGAGGAAG CTAGGCCAC TGCTCGGA AGTGCACCA 9930  
 T E K R H T R Q G I O T H L T L K E S T G D V L P K C D  
 CCCGCCACCA ACATAACCTC ACTCAGACCG GAAACAGCCG GTCTTTCCTP TTCTGCCCCAT TCCCACAGG TGAACAACTC ACTCTCAC 10080  
 P A A T \*  
 CTGACCCCTT TATCAGACTT CAATTATTA ATAAGAACG AGAGCTGAC AGTGTGACI GAAAACCTG AAAAAGCTT 10170  
 TTATCACCCCG CGCATGCC CGACTGGACA GATCCAGAAC GACCAAAGAT CACAAAGCT ACCAGCTCAC TGCTCACTC TCAACAACTC 10260  
 ATCACCACCT AACACATCA TATAAAATCA TAATAATCG AGCTTACCG TTAATCGA AAAACTCTP TTCTGCCCCAT TGCTGATTAAG 10350  
 AAATTAATT CACATCAATA CCTTCTCTCT TGAACTCTCT TGAGGTTTAT GAGGCGCTAA ~~GAGGCGCTAA~~ CCTTGCAATT TATGGTTAA 10440  
 TTTTAATTAA GTGAGTTAT CAAAGGAGA ACTTAGGAGA ACTCCAAATA CTCTCCATT GTCTGCTT GTGCTGATTAATACCCATT  
 TAGCCACATCG CGCATGAGC CTAGCTTCCG CTAGCTTCCG CTAGCTTCCG TGCTGCTCT TGCTGCTCT TGACTCTGCT 10530  
 ATCGGGTACG CGCTACTCG TACCAAGG CATCATAAA ~~CGCATACCGG~~ CAGGAGCTAA ACCAGCTCG CGCCGACCC CGCTGCTCG  
 AGCGAACTGAG CGACCTCTCT CTAGCTGTT TGCTCTCTCT AGTCTCTCT CATCAAGCTG CATAACGATA TGCTGCTG TGACTCTGCT 10620  
 TGCTCTGACT AGCTGACCGA CTGACCTCA ACCGAGAAAAA TGAGGAGATG GTAGTCTCAC GATTCCTAT AGAGCTACG ACTGAGCTGA  
 off314 N L T Q L  
 >DNA  
 AAAAAAAAT GGTACTGAGG TATCTACCC AACCCGTTA TTTCATCATG TGTTGAAAGA GAAACAGTA AAATGCTCTG GAAATGTAAGA 10710  
 K E K G Z E V B R A T A L P S F V Y R E K R P G H V K  
 AAATTCCTC TTTCCTGCTG GTGCTAAACAA AACACATCGA GAAACATCGA CAAGACGATT GGAATTAATAA AATTTTCTG AAAGGTATT 10800  
 K F V P L C G A N K H H G E P S A R R L E L I H F S E K V L  
 GAATAACTG CACTTTTCTG TTGCTGAACT AGTTTCTAA GAATTAAGCA CGGATGAGA ATCATTATCT GATAATTATG TACATATCGA 10890  
 H M C H F Y L A E L V F K E L S T D E E S L S O U L L D I E  
 AGCTGACTTA TCTAAATTAC CTGATCATAT TATCATTGTT TGAAAGTT ATTACATCTT CACGGACTT CGTCATCTG CATAACGCAA 10980  
 A D L S K L A D H I I I V L E S Y S E F T E L G A F A V E K  
 GCAATTACGC AAGAAATTAA TAATGTTAA CAATACAAA TTTCATTAATG AGAATACATT TATAAAATCG GGAATTAATAA AGGCTTATTAC 11070  
 Q L R K K E L I I V M H T E F I N E K S Y I N H G P I K A I T  
 TCAGCAATCA CAACAACTG CTGCTTCTCT ACATTATAAA ATGACAGAG TGATTGAAAG TATAGAGGCC TGCTGATGGA TTGGCGAAT 11160  
 Q Q S Q O S C H F L H Y E N T E G I E S I E R S D G I G E I  
 ATTCGACCCC CTATGATA TTCTCTCTA GAAACACAGA GCAATTCTCA GAACTTTAA AAAACAGAG TTGATCTCTT CCAGTAACCTT 11250  
 P P O L Y D I L E K H D R A I S E R T L K E E E L D P S B R Y  
 CAATAAAGAC TCAGTACGAT TTTCATGAGA CTAAATTTTT GTATGCTGCT CTTCGACT TAATGAACTC ATCGAAATTA TCAACAAAT 11340  
 N K O S V R F I N D O V I F V C G P L Q L H E L I K I I T K I  
 ATTTGGCACA GAAAGCCATT AACAAAAAAA TTCTCTAAAG CACTCTGTTA TTCTAATAGC TATTAGATAA ATATCATCGA CAATGGGAT 11430  
 P G T E S H Y K E N L L K H G I L I A R I I S C T N G I  
 TTATATCTCT TTGTATAAG ATATTTTAAAG TTATATGAC TTTCAGATG ACACATCTG ATCAATGTTT AAAGTTTTT TTCTCAAGAA 11520  
 Y Y S L Y Y Y P V K D F D I D M I S H N F K V F F L E N  
 CAAGCCAGAA AGGATGAGGG TATATGAGAA TATATAGCTT ATTGATTCT CACACATGCA TGACTAAGGG ATTGCTCTCT GAGTAATGC 11610  
 K P E R H R K V Y E N I \*  
 RT M R I Y S L I D S Q T L H T K G F A S E V N  
 GATCAGCTGA CGCGCCAAAG AACATGGGATA TAGCTAAAGAA AAAGAGGGT ATGACAAACAA TTTCATCCC GTCATCAAAA GTTAAATTAA 11700  
 R S P E P P K K K M D I A K K K G G H R T V B E V K L  
 TTCAATATG TTGATGAAAT AATGTTTTT CGAAGCTCCG ATTCGAAATG GTCTGATATO CATTGTTAA AAACCGATCA ATAAAAACCA 11790  
 I Q Y W L K H H H V B K L P H H A A Y A P V K H R S I K S  
 ATGCTTCTTAT ACATGCCAA TCAAAAGATA ATGTTATGAT GAAATAGAT CTCAACAGAT TTTCCTCTC ATAAATTTT ATGATTTTG 11880  
 H A L L H A E S K H Y V K I O L K D F F P S I K F T D F  
 AGTAGCCAT CACTCTTAT CGAGATGCA TTGATTTAC TACAAATGATA GAAAGGAGT TACTACACT TATAAAAGG ATCTGCTTTA 11970  
 E E A F T T R D R I E F T T E Y D K E L L O L I K T I C F  
 TATCAGATG CACTCTCCCT ATCGGGTTTC CTACATCTCC ATTAAATGCA ACCTTTGCG CAAGACACT TGATGAAAAA CTGACGCAA 12060  
 I S D S T L P I G F P T S P L I A H P V A R E L D E K L T Q  
 AACATGATAA ATGTTATGAA TTAAATGCA CTTATACCC ATATGCTGAT GATATTTATG TTCTCTACAA TATGAAAGGG GCTGACCAAT 12150  
 K L N A I D K L H A T T Y R A D D I I V S T H N E G A S E  
 TAATCTGGA TTGTTTAAAG AGAACATGAA AGAGATGG TTCAAGCTT AAAATTACAA TTAAATTAATGATTTTG AGTCTCTGG 12240  
 L I I D C F E R T H K I I G P D F K T B I K E F K I C S A S  
 GAGGAAGTAT AGTAGTTACG GGATGAAAG TTGCGACCA TTTCATATT ACATACATA GATCAATGAA AGATAAAATA AGATGCTAC 12330  
 G G S I V V T G L K V C H O F H I T L H R S H K D K I R L H  
 TTCTCTCTT ATCAAAGGGC ATATTTAAAG ATGAGATGCA TAATAACCT TTCTGTTATA TTGCTTATG AAAGATATA GACCCCTATT 12420  
 L S U L L B K O I L K D E O H N K L S G T I A Y A K D I D P H  
 TTATACAA AGTGAACACA AAATTTTTC AGAGATATAA ATGAGATGCA ATGCTCACCA ACAGAGTGA ATAAACTTAA TATTTGGAT 12510  
 P Y T K L H R K Y F Q E I K W I Q K L H H K V E \*  
 GCACCCCAAT AACCTCATGTT ATTAAATGG GAAACATATAA GCGCTTCTG GATGACTAC ACTCTAGAGA ATGTTATAC AAAAGTGTAT 12600  
 AACATTTTTT CAACCTATAA TAAATACAG CAAACATCA GCAATGGGGG CATTTCACG ~~CCTCTGCTG~~ CCTCCGCAA AACCTGCT 12688

FIGURE 17

(A)



-371 TGGCATCTATTAAGAAGGTTAGGAAGAAAATAAAGTATCAAAGATATTGAAATATAT  
 -311 TATACGCAGAGCGTTCTATGCCCTGTATCTATTTACTGGATAGTGTCAACTACCGCAC  
 -251 ACTGTGTGAACTAGCTTTAAGCGATAAAGCAAGATGATGTTTATCTAAAAATTATTGT  
 -191 TAGATCCGTTGTTCTCGCTAATAAAATGAAACGAAAATACTCAATGACTGATGGTTA  
 -131 TCAGGTCACTGCTTGGGGCTAGCTATGTTAGGAGCGCTTTGATAGAAAAGACACTTGA  
 -71 CCGATTGCGGCTTGAGATTATGAATTGAAAACCGTAGAAAATCAACATTAACATATGA  
 +1  
 -11 TAAGATTCCGTATGCGCACCCCTAGCGAGAGGTTTATCATTAAAGGTCAACCTCTGGATGT  
 IR ----->  
 49 TGTTCGGCATCCGATTGAATCTGAGTTACTGTCCTGTTCTGTTGAAACGGAGAG  
 <-----  
 109 CATGCCCTGATGCTCTCGAGCCAACCCAGGAAACCGTTTTCTGACGTAAAGGTGCGC  
 ssDNA-----|-----  
 169 AACTTCATGAAATCCGCTGAATATTGAAACACTTTAGATTGAGAAAATCTCGGCCCTACC  
 - IR MetLysSerAlaGluTyrLeuAsnThrPheArgLeuArgAsnLeuGlyLeuPr  
 229 TGTATGAAACAATTGATGACATGTCTAAGGCAGTCGCATATCTGTTGAAAACACTCG  
 ovalMetAsnAsnLeuHisAspMetSerLysAlaThrArgIleSerValGluThrLeuAl  
 289 GTTCTTAATCTATACAGCTGATTTGCTATAGGATCTACACTGTAGAAAAGAAAAGGCC  
 gLeuLeuIleTyrThrAlaAspPheArgTyrArgIleTyrThrValGluLysLysGlyPr  
 349 AGAGAAGAGAATGAGACCATTTACCAACCTCTCGAGAACCTTAAGCCTACAGGATG  
 oGluLysArgMetArgThrIleTyrGlnProSerArgGluLeuLysAlaLeuGlyTr  
 409 GGTCTACGTAACATTGATGATCTGCTATCTCTTTCTATTGATTTGAAAA  
 pValLeuArgAsnIleLeuAspLysLeuSerSerProPheSerIleGlyPheGluLy  
 469 GCACCAATCTATTGAAATACTGCTACCCCGCATATTGGGCAACTTATACTGAAATAT  
 sHisGlnSerIleLeuAsnAsnAlaThrProHisIleGlyIlaAsnPhelleLeuAsnII  
 529 TGATTGGAGGATTTTCCAAAGTTAACGCTAACAAAAGTTTGGAGTGTCCATTG  
 eAspLeuGluAspPhePheProSerLeuThrAlaAsnLysValPheGlyValPheHisSe  
 589 TCTGGTTATAATCGACTAAATCTTCAGTTTGACAAAAATATGTTGTTATAAAAATCT  
 rLeuGlyTyrAsnArgLeuIleSerSerValLeuThrLysIleCysTyrLysAsnLe  
 649 GCTACCACAAAGTGTCTCCATCATCACCTAAATTAGCTAATCTAAATGTTCTAAACTTGA  
 uLeuProGlnGlyAlaProSerSerProLysLeuAlaAsnLeuIleCysSerLysLeuAs  
 709 TTATCGTATTCAAGGTTATCGAGTAGTCGGGCTTGATATACGAGATATGCCGATGA  
 pTyrArgIleGlnGlyTyrAlaGlySerArgGlyLeuIleTyrThrArgTyrAlaAspAs  
 769 TCTCACCTTATCTGACAGCTATGAAAAAGGTTAAAGCAGTGTGATTTTTATTTTC  
 pLeuThrLeuSerAlaGlnSerMetLysLysValValLysAlaArgAspPheLeuPheSe  
 829 TATAATCCCAGTGAAGGATTGGTTATTAACTCAAAACCTTGATTTAGTGGGCTCG  
 rIleIleProSerGlyLeuValIleAsnSerLysThrCysIleSerGlyProAl  
 889 TAGTCAGAGGAAGTACAGGTTAGTTATTCAACAGAGAAAAGTGGGATAGCTAGAGA  
 qSerGlnArgLysValThrGlyLeuValIleSerGlnGluLysValGlyIleGlyArgG  
 949 AAAATATAAAGAAATTAGACAAAGATACATCATATAATTGCGGTAAGTCTTCTGAGAT  
 uLysTyrLysGluIleArgAlaLysIleHisIlePheCysGlyLysSerSerGluIle  
 1009 AGAACACGTTAGGGATGGTTGTCAATTAACTGTTAAGTGTGGATTCAAAAGCCATAGGAG  
 eGluHisValArgGlyTrpLeuSerPheIleLeuSerValAspSerLysSerHisArgAr  
 1069 ATTAATAACTTATATTAGCAAAATTAGAAAAAAATATGAAAAGAACCCCTTAAATAAAAGC  
 gLeuIleThrTycIleSerLysLeuGluLysLysTyrGlyLysAsnProLeuAsnLysAl  
 1129 GAAGACCTAATGGCTTCGTTTAAACCTAAAGCTCATAGGTTGAAAATTGAGCACTTC  
 aLysThr  
 1189 TTCGTCCAACCAAGTTATTAGTTCCTGCAATCGTTCTGAG

FIGURE 18

Oligo 2337

tcaccctgaaagac~~t~~gatt~~g~~c~~t~~ttac~~c~~tg~~a~~agagaagccgaaatggc~~g~~gaacatctgg 60  
 cg~~c~~cg~~g~~tt~~a~~agg~~c~~ctatcg~~g~~aa~~g~~ag~~t~~tc~~g~~gc~~t~~ttaaaATATGGCCTCTGCAGGTTT 120  
 TTGCTGTCGCA~~CG~~TGATGGC~~G~~TTCAAG~~A~~TATCGTGT~~T~~ATCTGCTT~~T~~GCCAGCAGTG RNA~~A~~ a2 180  
 AACGACACGCGTCGACTACCGGAAGTTCTATGCACAATTAGACGAAAGCGGT~~C~~GT~~C~~AC 240  
 CGCAATAGCGTTCCGGCTTTGTGCCCCGAGGGTCGGCAGTCGCTGACTTAACGCCAG 300  
 CGTTATCGCAAAAGGCCAAAACCGCCCTCCAGCCGCTAGCGACTGAATTGGGTC 360  
 TAGTATGTCCATATAACCCAAGTCGCTTCATTG~~T~~ACCTGAGTACGCTTCGGTACGT~~CG~~CG 420  
 ATCATACAGGTATATGGGTTT~~C~~AGCGAAGTAA~~C~~ATGGACTCATGCCAACGCCATGCCAGCG  
 GC~~G~~GC~~C~~CTTTGGCCGCCGT~~G~~T~~G~~TGGAGAGT~~G~~GAATGGATGGT~~C~~ACCCGGACAACCC~~T~~  
 M D A T R T T L L 480  
 TGGCGCTCGATTGTT~~C~~GGCTCGCCGGGCTGGAGCGCGATAAGAAATACAGCGACTGC  
 A L D L F G S P G W S A D K E I Q R L H 540  
 ATGCGCTCAGTAATCATGCCGGACGCCATTACCGACGCATTATTCTTCTAAACGCCACG  
 A L S N H A G R H Y R R I I L S K R H G 600  
 GTGGTCAGCGGCTGGT~~G~~T~~G~~TAGCCC~~T~~GATTACTTGCTAAAACCGTACAGCGAACATT~~C~~  
 G Q R L V L A P D Y L L K T V Q R N I L 660  
 TTAAGAACGTC~~C~~TTTCACAATTCCGCTTCCCTTTGCTACAGCCTACCGACCAGGTT  
 K N V L S Q F P L S P F A T A Y R P G C 720  
 GCCCAATCGT~~C~~AGCAACGCCAGCCACACTG~~C~~AAACAGCGCAGATCTGAAACTCGATA  
 P I V S N A Q P H C Q Q P Q I L K L D I 780  
 TCGAAAAC~~T~~TTTCGATAGCATTAGCTGGT~~T~~ACAGGT~~T~~GGCTG~~T~~GT~~T~~TCGCCAGGCC  
 E N F F D S I S W L Q V W R V F R Q A Q 840  
 AGTTGCCACGT~~A~~TGTGGT~~A~~CCATG~~T~~GT~~C~~ACCTGGATT~~T~~GT~~T~~TATAACGACGCC~~T~~TAC  
 L P R N V V T M L T W I C C Y N D A L P 900  
 CGCAGGGGGCACCAACTCGCCAGCCATTCCAA~~T~~TGT~~G~~T~~A~~GGCTGGCTTTGATGAAC  
 Q G A P T S P A I S N L V M R R F D E R 960  
 GCATAGGGGAATGGT~~G~~T~~C~~AGGCTCGGGAAATTACCTACACCCG~~T~~ACTGCC~~T~~GACATGA  
 I G E W C Q A R G I T Y T R Y C D D M T 1020  
 CCTTTTCAGGT~~C~~ACTTCATGCCGCCAGGTAAAAAAGTGTGCCGATTGTTAGCGG  
 F S G H F N A R Q V K N K V C G L L A E 1080  
 AGCTGGGCTGAGCCTCAATAACGCAAAGGCTGCC~~T~~GATAGCTGCC~~T~~GTAAGCGCCAGC  
 L G L S L N K R K G C L I A A C K R Q Q 1140  
 AAGTAACCGGATTGTTGTTAATCACAAGGCCACAGCTGCCG~~T~~GAA~~G~~CGCCGGCG  
 V T G I V V N H K P Q L A R E A R R A L 1200  
 TCGCTCAGGAGGT~~G~~CATTG~~C~~AAAAATGGCTTATTCG~~C~~ATCTTAGTCATCG~~T~~G  
 R Q E V H L C Q K Y G V I S H L S H R G 1260  
 GTGAAC~~T~~TGATCCTCTGGCGATCTCCACG~~C~~ACAGGCAACGGCGTATCTTATGCTT~~G~~  
 E L D P S G D L H A Q A T A Y L Y A L Q 1320  
 AGGGAGAATAAAACTGGTATTG~~C~~AA~~T~~CAACCC~~T~~GAGGT~~G~~AGGCC~~T~~CAACAGCGA  
 G R I N W L L Q I N P E D E A F Q Q A R 1380  
 GAGAGAGT~~G~~TAAAGCGAATG~~T~~GGTTG~~C~~ATGGTAAGAAAAGCGTCAGGCAGACGTT~~T~~CTG  
 E S V K R M L V A W \* 1440  
 CCTGACCGTTAGGGAGAattactgcaactgcgcggcaattagcgccagcggcgtca 1500  
 aaatcatccgtcgccgtat~~t~~aaactcgctcg~~g~~gacaaaacgtgacagcatac~~t~~ca  
 cagaaggccaggatctggctgccagcagg~~ttt~~catcg 1540  
 Oligo 2336

FIGURE 19

2' DEO = DEO-TEO-GOE

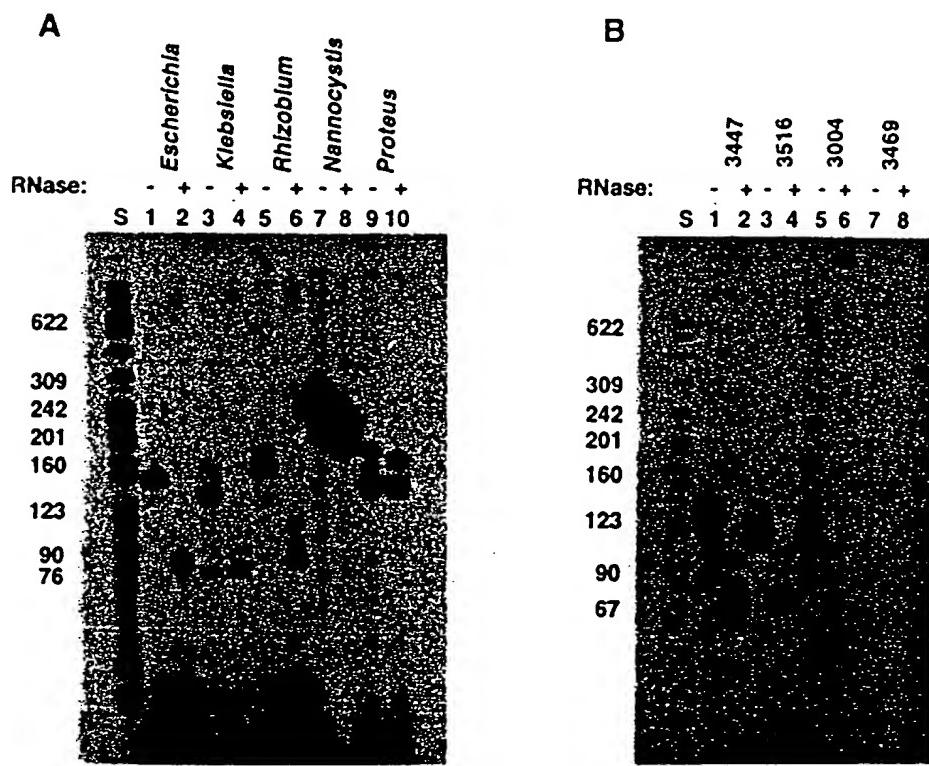


FIGURE 20

## RHIZOBIAL ISOLATES

Strain (legume host genus)	USDA strain no.	Geographic source (date)	msDNA produced <sup>b</sup>
<i>Rhizobium</i> sp. ( <i>Acacia</i> )	3002	Brazil (1959)	+
	3003	Africa (1950)	
	3325	Morocco (1974)	
	3838	? (1976)	
<i>Bradyrhizobium</i> sp. ( <i>Aeschynomene</i> )	3516	Florida (1972)	+
	4362		+
<i>Bradyrhizobium</i> sp. ( <i>Albizia</i> )	3004	Maryland (1952)	+
<i>Bradyrhizobium</i> sp. ( <i>Apios</i> )	3240	Maryland (1939)	
<i>Bradyrhizobium</i> sp. ( <i>Arachis</i> )	3339	Thailand (1979)	
	3341	Hawaii (1978)	
<i>Rhizobium</i> sp. ( <i>Astragalus</i> )	3854	Alaska (1962)	
<i>Rhizobium</i> sp. ( <i>Cajanus</i> )	3472		
<i>Bradyrhizobium</i> sp. ( <i>Canavalia</i> )	3317	Brazil (1974)	
<i>Rhizobium</i> sp. ( <i>Cicer</i> )	3378		
	3379	Mexico (1963)	
<i>Bradyrhizobium</i> sp. ( <i>Coronilla</i> )	3165	Virginia (1935)	
	3167	? (1961)	
<i>Bradyrhizobium</i> sp. ( <i>Crotalaria</i> )	3384	Brazil (1967)	
<i>Bradyrhizobium</i> sp. ( <i>Desmodium</i> )	3225	Ecuador (1948)	
<i>Bradyrhizobium</i> sp. ( <i>Erythrina</i> )	3241		
	3242	Maryland (1939)	+
<i>Rhizobium fredii</i>	191	China (1979)	
<i>Rhizobium leguminosarum</i>	2370	Illinois (1933)	
	2429	Hawaii (1978)	
	2435	Holland (1955)	
	2480	Tennessee (1951)	
	2489		
<i>Rhizobium</i> sp. ( <i>Lens</i> )	2426		
	3404	Colombia (1979)	
<i>Rhizobium loti</i>	3084	Maryland (1946)	+
	3468	New Zealand (1961)	
	3469		+
	3471		
	3503		
<i>Bradyrhizobium</i> sp. ( <i>Lotus</i> )	3669	California (1968)	
	3074	Minnesota (1954)	
	3470	California (1916)	
<i>Rhizobium</i> sp. ( <i>Lupinus</i> )	3040	Florida (1940)	
<i>Bradyrhizobium</i> sp. ( <i>Lupinus</i> )	3045	Florida (1946)	
<i>Bradyrhizobium</i> sp. ( <i>Macrotyloma</i> )	3451	Zimbabwe (1960)	
<i>Rhizobium medicago</i>	1097	North Dakota (1948)	
<i>Rhizobium meliloti</i>	1011	Maryland (1933)	
<i>Rhizobium phaseoli</i>	1021a	North Dakota (1948)	
	2667	Washington (1948)	
	2669		
	2674	Brazil (?)	
	2676	Colombia (1972)	
	3256	Illinois (1941)	
<i>Rhizobium</i> sp. ( <i>Robinia</i> )	3436		
<i>Bradyrhizobium</i> sp. ( <i>Stylosanthes</i> )	3441	Brazil (?)	
	3477	Colombia (1976)	
<i>Rhizobium trifolii</i>	2046	Virginia (1934)	+
	2048	Illinois (1934)	
	2063	Florida (1939)	
	2065	Alabama (1952)	
	2116	South Carolina (1944)	
	2134	? (1974)	
	2145		
<i>Rhizobium</i> sp. ( <i>Trigonella</i> )	2156	California (1920)	
<i>Rhizobium tropici</i>	1177	Florida (1939)	
<i>Bradyrhizobium</i> sp. ( <i>Vigna</i> )	2744	Brazil (?)	
	3447	Thailand (1979)	+
	3456	Wisconsin (1966)	

<sup>a</sup> All strains are from the USDA Beltsville Rhizobium Culture Collection, provided by Peter van Berkum.

<sup>b</sup> As defined by detection of radiolabeled msDNA by the RT extension method.